Some data collection and analysis of the distribution of Champion Primes for non-CM Elliptic Curves

---

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Mathematical Sciences

---

by
Alan R. Hahn
August 2018

Accepted by:
Dr. Kevin James, Committee Chair
Dr. William Bridges
Dr. Hui Xue

Abstract

For a fixed non-singular elliptic curve E given by $y^2 + axy + cy = x^3 + bx^2 + dx + e$, the frequency of extremal primes for E up to a given X value is of interest, where an extremal prime p is a prime for which the order of E defined over $\mathbb{F}_p$ is a maximum or minimum with respect to Hasse's Theorem. For CM elliptic curves this distribution is known to not be curve dependent, and in this paper some preliminary work on determining the distribution of such primes for the non - CM case is presented.

Acknowledgments

I'd like to give sincere thanks to Dr. James for helping me learn about elliptic curves and helping me with my project. I'd like to thank Dr. Bridges for helping me learn some data analysis and helping me use the data analysis for this project. I'd like to thank Dr. James, Dr. Bridges, and Dr. Xue for being on my committee and for helping me prepare my thesis. I'd also like to thank the Palmetto Cluster on Clemson's campus for allowing me to use their infrastructure to conduct some of my research.

Table of Contents

List of Tables

List of Figures

Chapter 1

Elliptic Curves and Motivation

1.1 Introduction to Elliptic Curves

A plane curve of degree n over a field K is the set of $(x, y) \in K \times K$ which satisfy an implicit polynomial of degree n in x, y with coefficients in K;

$$f(x, y) = \sum_{i,j}^{n} a_{i,j} x^i y^j = 0, \text{where } i + j \leq n.$$

The solutions to such an equation are called affine solutions.

The homogenization of a polynomial $f(x, y)$ of degree n is defined to be

$$F(x, y, z) = \sum_{i,j}^{n} a_{i,j} x^i y^j z^{n-i-j}$$

so that each monomial in the sum is of the same degree. This homogenized polynomial has the property that for any $(x, y, z) \in K \times K \times K$ such that $F(x, y, z) = 0$, $F(tx, ty, tz) = 0$, for all $t \in K$. In this setting the solutions are then identified, so that a solution to $F(x, y, z)$ is

$$[x, y, z] = \{(x', y', z') : x' = tx, y' = ty, z' = tz, 0 \neq t \in K\}.$$

Such a solution to an equation $F(x, y, z) = 0$ is called a projective solution, and the set of such solutions to an equation with a polynomial of degree n, $F(x, y, z) = 0$, is called a projective plane

curve of degree n.

For an affine solution (x, y) to $f(x, y)$, $[x, y, 1]$ is a projective solution to the homogenized polynomial $F(x, y, z)$, and similarly, for a projective solution $[x, y, z]$ with $z \neq 0$ to $F(x, y, z)$, $(x/z, y/z)$ is an affine solution to $f(x, y)$. The projective solutions $[x, y, z]$ such that $z = 0$ are called points on the curve at infinity.

A plane curve is called singular at a point $p \in K \times K$ on the curve if $\frac{\partial f}{\partial x}(p) = \frac{\partial f}{\partial y}(p) = 0$, and is called non-singular if there are no points at which the curve is singular. Every cubic plane curve $f(x, y)$ is isomorphic to a curve given by

$$y^2 + axy + cy = x^3 + bx^2 + dx + e.$$

If the characteristic of K is not 2 then every cubic plane curve is isomorphic to a curve given by

$$y^2 = f(x) = x^3 + gx^2 + hx + i,$$

where $f(x)$ has discriminant $-4g^3i + g^2h^2 + 18ghi - 4h^3 - 27i^2$ [7]. Checking whether a curve is non-singular is equivalent to checking whether the discriminant of $f(x)$ is 0, which is equivalent to checking whether $f(x)$ has distinct roots over the algebraic closure of K.

Definition 1. An elliptic curve E over a field K is a non-singular projective cubic curve which has an element in $K \times K \times K$.

As stated above, an elliptic curve is isomorphic to a curve given by $y^2 + axy + cy = x^3 + bx^2 + dx + e$ and in such a case the elliptic curve is denoted $E_{a,b,c,d,e}$. If the characteristic of the field is not 2 then the elliptic curve is denoted $E_{g,h,i}$ and in the case that $g = 0$, $E_{h,i}$. When the context is clear, E will denote an elliptic curve. An elliptic curve has one point at infinity, denoted O, and $E(K)$ denotes the points of E which have coordinates in K along with the point at infinity, that is $E(K) = \{(x, y) : f(x, y) = 0,\ x, y \in K\} \cup \{O\}$.

A group structure may be defined on an elliptic curve along with a specified projective solution acting as the identity element. $E(K)$ is an abelian group, and $\#E(K)$ denotes the number of points of $E(K)$. The following three theorems are related to the structure of $E(\mathbb{Q})$ are useful to know. Further information on these theorems and the structure of $E(\mathbb{Q})$ may be found in [7].

Theorem 2. (Mordell) Let E be given by $y^2 = x^3 + ax + b$, with $a, b \in \mathbb{Z}$. Then $E(\mathbb{Q})$ is a finitely

generated abelian group.

Theorem 3. (Nagell - Lutz) Let E be given by $y^2 = f(x) = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{Z}$, let D be the discriminant of $f(x)$, and let $(x, y) \in E(\mathbb{Q})$ be a point of finite order. Then $x, y \in \mathbb{Z}$ and either $y = 0$ or $y \mid D$.

Theorem 4. (Mazur) Let E be defined over Q and suppose that $E(\mathbb{Q})$ has a point of finite order. Then the set of points of finite order in $E(\mathbb{Q})$ forms a subgroup that has one of the following forms:

1. A cyclic group of order N with $1 \leq N \leq 10$ or $N = 12$.

2. The product of a cyclic group of order two and a cyclic group of order $2N$ with $1 \leq N \leq 4$.

The attention now turns to $E(\mathbb{F}_p)$. For an elliptic curve E defined over $\mathbb{F}_p$, $\#E(\mathbb{F}_p)$ will be finite as $\mathbb{F}_p \times \mathbb{F}_p$ is finite, and a bound on $\#E(\mathbb{F}_p)$ is known, due to Hasse:

Theorem 5. (Hasse)

$$|(p+1) - \#E(\mathbb{F}_p)| \leq 2\sqrt{p}.$$

The quantity $(p+1) - \#E(\mathbb{F}_p)$ is of special interest and is denoted $a_p = (p+1) - \#E(\mathbb{F}_p)$. The bound in Hasse's theorem is best possible in the sense that for any integer u within the bound, there exists an elliptic curve with $a_p = u$; this is a corollary of Deuring's theorem [2].

Definition 6. An extremal prime p is a prime for which $a_p = \pm\lfloor 2\sqrt{p} \rfloor$, i.e. a prime for which $\#E(\mathbb{F}_p)$ is a maximum or minimum with respect to the Hasse bound.

Definition 7. A champion prime is an extremal prime for which $a_p = -\lfloor 2\sqrt{p} \rfloor$, i.e. a prime for which $\#E(\mathbb{F}_p)$ is a maximum with respect to the Hasse bound. Similarly, a trailing prime is an extremal prime for which $a_p = \lfloor 2\sqrt{p} \rfloor$, i.e. a prime for which $\#E(\mathbb{F}_p)$ is a minimum with respect to the Hasse bound.

Let E be defined over $\mathbb{C}$. An algebraic endomorphism is a homomorphism from E to itself defined by rational functions, that is, an algebraic endomorphism is a non - trivial homomorphism $\phi : E(\mathbb{C}) \to E(\mathbb{C})$ such that

$$\phi(x, y) = \left(\frac{\text{polynomial in } x, y}{\text{polynomial in } x, y}, \frac{\text{polynomial in } x, y}{\text{polynomial in } x, y} \right).$$

Throughout, algebraic endomorphisms will be called endomorphisms. The set of all endomorphisms of a curve is a ring under composition of functions and point-wise addition, denoted $End(E)$. The multiplication - by - n maps, $[n] : E(\mathbb{C}) \to E(\mathbb{C})$ defined by $P \to nP = \sum_{i=1}^{n} P$, where addition is with respect to the group law, are endomorphisms and by identifying $[n]$ with n, the set of endomorphisms of E contains $\mathbb{Z}$, that is that $End(E) \supseteq \mathbb{Z}$. A curve E for which $End(E) \neq \mathbb{Z}$ is said to have complex multiplication, and the curves with complex multiplication are density 0 in the set of all elliptic curves [4].

As noted above, for E defined over K with characteristic not equal to 2 or 3, E is isomorphic to a curve given by $y^2 = x^3 + ax + b$. Suppose $u \in \overline{K}^*$ and multiply through by u^6 to obtain $u^6y^2 = u^6x^3 + u^6ax + u^6b$. Taking as a change of variables $X = u^2x$, $Y = u^3y$ gives another equation $(u^3y)^2 = (u^2x)^3 + au^4(u^2x) + u^6b$, or $Y^2 = X^3 + au^4X + u^6b$. Note that this defines an isomorphism between the curve E given above and the curve E' given by $Y^2 = X^3 + au^4X + u^6b$. Two isomorphic curves E and E' are said to be different models for the same elliptic curve. In the case that E is defined over $\mathbb{Q}$, and taking u to be a prime p which does not divide the discriminant $4a^3+27b^3$ of E, note that in one model $p \nmid 4a^3+27b^3$ but p divides the discriminant $4a^3p^{12}+27b^3p^{18}$ of E'. An elliptic curve E is said to be minimal at p if the power of p dividing the discriminant of E is a minimum over all models of E. A minimal model is a model of an elliptic curve which is minimal at all primes p. With this in mind it is useful to note that the above definition for $E(K)$ is taken to mean $E(K)$ for some minimal model of E.

1.2 Background Information and Motivation

The goal of this paper is to investigate $\#\{p \leq x : E \text{ over } \mathbb{F}_p \text{ is non-singular, } a_p = -\lfloor 2\sqrt{p} \rfloor\} = \pi_E^{\text{Champ}}(x)$, the distribution of champion primes of a given elliptic curve E over $\mathbb{F}_p$. This is a refinement of the Sato - Tate conjecture, which is a theorem in the case that E is defined over K a totally real field, due to Taylor, Clozel, Harris, Shepherd - Barron [3];

Theorem 8. (Sato-Tate) Suppose E is an elliptic curve over $\mathbb{Q}$ which does not have complex multiplication. For $-1 \leq \alpha \leq \beta \leq 1$, $x > 1$, as $x \to \infty$,

$$\#\{p \leq x : \alpha \leq \frac{a_p}{2\sqrt{p}} \leq \beta\} \sim \frac{2}{3\pi}\Big(\int_{\alpha}^{\beta} \sqrt{1-t^2}\, dt\Big)\, \frac{x}{\log(x)}.$$

The Sato - Tate conjecture is a statement about the distribution of primes which fall within an interval within the Hasse bound, while the distribution investigated in this paper is the distribution of primes which achieve the extremes of the Hasse bound. This question has been partially answered. In the case that the curve has complex multiplication the result is known, due to James and Pollack [5]:

Theorem 9. (James, Pollack) For E an elliptic curve defined over $\mathbb{F}_p$ which has complex multiplication, as $x \to \infty$,

$$\pi_E^{\text{Champ}}(x) \sim \frac{2x^{3/4}}{3\pi \log(x)}.$$

The same result holds for trailing primes. For the case in which the curve does not have complex multpilication, an average result is provided by James and Giberson [3]:

Theorem 10. (James, Giberson) For A, $B > x^{3/4}\log(x)$ with $AB > x^{7/4}\log(x)$, as $x \to \infty$,

$$\frac{1}{4AB} \sum_{\substack{|a| \leq A \\ |b| \leq B}} \pi_{E_{a,b}}^{\text{Champ}}(x) \sim \frac{8x^{1/4}}{3\pi \log(x)}.$$

It is noteworthy that in the previous theorem the effect by CM curves is negligible as CM curves are density 0 in the family of all elliptic curves [4]. It is also useful to note that as $x \to \infty$,

$$\frac{2}{3\pi} \int_2^x \frac{dt}{t^{3/4}\log t} \sim \frac{8x^{1/4}}{3\pi \log(x)}.$$

As stated, Theorem 10 is an on average result and a goal of this paper is to provide evidence for a non - average conjecture in the case that the curve does not have complex multiplication. In the case that a given elliptic curve has complex multiplication, the distribution of champion primes is not curve dependent; a goal of this paper is to obtain data to help determine whether the distribution of champion primes is curve dependent for the non - CM case.

1.3 Computing a_p

Given an elliptic curve E defined over $\mathbb{F}_p$, recall that

$$a_p = (p+1) - \#E(\mathbb{F}_p)$$

so that knowing $\#E(\mathbb{F}_p)$ is equivalent to knowing a_p. Thus in order to determine whether a given prime p is a champion prime for E, i.e. whether $a_p = -\lfloor 2\sqrt{p} \rfloor$, it suffices to know $\#E(\mathbb{F}_p)$. A naive method for computing $\#E(\mathbb{F}_p)$ is given below. Throughout, p is an odd prime.

Definition 11. $a \in \mathbb{F}_p$ is called a quadratic residue modulo p if there exists $0 \neq x \in \mathbb{F}_p$ such that $a \equiv x^2 \pmod{p}$, and a is called a quadratic non - residue otherwise.

Definition 12.

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } a \equiv 0 \pmod{p} \\ 1 & \text{if } a \text{ is a quadratic residue modulo } p \\ -1 & \text{if } a \text{ is a quadratic non - residue module } p \end{cases}$$

where $\left(\frac{a}{p}\right)$ is called the Legendre symbol.

Consider $y^2 = f(x) = x^3 + gx^2 + hx + i$ defined over $\mathbb{F}_p$ and note that for $x_0 \in \mathbb{F}_p$, if $f(x_0)$ is a quadratic residue then the above equation is satisfied for those $y \in \mathbb{F}_p$ which square to $f(x_0)$. In this case there are two values y_0 and $-y_0$ in $\mathbb{F}_p$ which square to $f(x)$ so that (x_0, y_0), $(x_0, -y_0)$ are the two points on $E(\mathbb{F}_p)$ which correspond to x_0. If x_0 is a quadratic non - residue, the above equation is not satisfied by any $y \in \mathbb{F}_p$ so that for x_0 a quadratic non - residue, there are no points on $E(\mathbb{F}_p)$ which correspond to x_0. For $f(x_0) = 0$, the equation above is satisfied for $y = 0$ so that $(x_0, 0)$ is the point on $E(\mathbb{F}_p)$ which corresponds to x_0. Thus

$$1 + \left(\frac{f(x)}{p}\right) = \begin{cases} 1 & \text{if } f(x) \equiv 0 \pmod{p} \\ 2 & \text{if } f(x) \text{ is a quadratic residue modulo } p \\ 0 & \text{if } f(x) \text{ is a quadratic non - residue module } p \end{cases}$$

counts the number of points on $E(\mathbb{F}_p)$ corresponding to $x \in \mathbb{F}_p$ so that summing over all $x \in \mathbb{F}_p$ and including the point at infinity gives $\#E(\mathbb{F}_p)$, that

$$1 + \sum_{i=0}^{p-1} \left(1 + \left(\frac{f(i)}{p}\right)\right) = (p+1) + \sum_{i=0}^{p-1} \left(\frac{f(i)}{p}\right) = \#E(\mathbb{F}_p).$$

Thus

$$a_p = (p+1) - \#E(\mathbb{F}_p) = -\sum_{i=0}^{p-1} \left(\frac{f(i)}{p}\right).$$

$\left(\frac{a}{p}\right)$ is fairly easy to compute and there exist efficient algorithms of doing so; an interested reader may see [1] for further information.

The algorithm outlined above is not very efficient, the algorithm runs in $O(p\log(p))$ time, and is listed here so the reader may see that calculating $\#E(\mathbb{F}_p)$ is a tractable problem. Indeed, there exist much more efficient algorithms for computing $\#E(\mathbb{F}_p)$ than the one listed above, particularly the Shanks - Mestre algorithm and Schoof's algorithm [1]. An overview of the Shanks - Mestre algorithm is given below and the algorithm itself may be found in Appendix B. Before discussing the Shanks - Mestre algorithm, a little more background is first needed.

Theorem 13. (Bezout) A projective plane curve of degree n and a projective plane curve of degree m intersect in $m * n$ points.

Consider an elliptic curve E and consider $E(\mathbb{C})$. Note that if a line intersects the cubic curve defining E, then by Bezout's theorem the line will intersect E in three places. If two of these intersection points are rational then so is the third, so that there is a way to obtain a third point $a * b \in E(\mathbb{Q})$ given two points a, $b \in E(\mathbb{Q})$. If a fourth point O is considered, a group law may be obtained; define

$$a + b = O * (a * b).$$

It is fairly straightforward to see that this is an abelian group with identity O, and the details are left to [7]. Now in order to obtain an explicit formula for the group law, it is useful to take the identity to be the point on E at infinity and to consider a, b as coordinates $a = (x_1, y_1)$, $b = (x_2, y_2)$. Letting

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1}$$

and

$$\nu = y_1 - \lambda x_1 = y_2 - \lambda x_2,$$

then the line joining a, b is given by

$$y = \lambda x + \nu.$$

Then plugging in $\lambda x + \nu$ into the equation for the elliptic curve,

$$y^2 = (\lambda x + \nu)^2 = x^3 + gx^2 + hx + i$$

gives a cubic equation in the one variable x,

$$0 = x^3 + (a - \lambda^2)x^2 + (b - 2\lambda\nu)x + (c - \nu^2).$$

Note that the roots of this equation are the three x_i coordinates of the intersection of $\lambda x + \nu$ with $y^2 = x^3 + gx^2 + hx + i$ and thus

$$0 = x^3 + (a - \lambda^2)x^2 + (b - 2\lambda\nu)x + (c - \nu^2) = (x - x_1)(x - x_2)(x - x_3).$$

Noting that the sum of the roots is the negative of the coefficient on the x^2 term gives

$$a - \lambda^2 = -x_1 - x_2 - x_3,$$

that

$$x_3 = \lambda^2 - a - x_1 - x_2.$$

Thus $y_3 = \lambda x_3 + \nu$ and thus

$$a + b = O * (\lambda^2 - a - x_1 - x_2, \lambda x_3 + \nu) = (\lambda^2 - a - x_1 - x_2, -(\lambda x_3 + \nu))$$

as the point at infinity on E corresponds to the point associated to vertical lines [7]. For the case that $a = b$, take

$$\lambda = \left.\frac{dy}{dx}\right|_a = \frac{f'(x_1)}{2y_1}$$

in the method discussed above.

Now that there is an explicit formula for the group law, algorithms which require the ability to use the composition law within the group may be implemented, including the Shanks - Mestre algorithm. The Shanks - Mestre algorithm consists of two parts - an algorithm called the Shanks Baby - Step Giant - Step method and an algorithm which uses an idea of Mestre. The Shanks Baby - Step Giant - Step method is implemented as follows:

Let G be a finite abelian group and let $g \in G$. Further, suppose $|G| \leq B$ and let $q = \lceil \sqrt{B} \rceil$, let

$$X = \{1, g, g^2, ..., g^{q-1}\},$$

and set $g_1 = g^{-q}$. Then if the order n of g is written

$$n = aq + r, \text{ with } 0 \leq r < q,$$

then $a \leq q$ as well by the choice of q. Thus for $a = 1, 2, ..., q$, the element g_1^a is computed and is checked to see whether $g_1^a \in X$. If $g_1^a \in X$, then

$$g_1^a = g^{-aq} = g^j, \text{ for some } 0 \leq j < q,$$

that

$$1 = g^{aq+j}$$

and thus n divides $aq + j$ so that n may be deduced by factoring $aq + j$.

Searching q times through the q elements of X will have running time $O(q^2)$ and will dominate the running time of this algorithm; a way to avoid this is to first sort X using an $O(q \ln q)$ method. Then searching through this sorted list will take only $O(\ln q)$ comparisons, bringing the total time to $O(q \ln q)$.

The Shanks - Mestre algorithm works by exploiting the group structure of E. Recall Hasse's theorem for $E(\mathbb{F}_p)$;

$$(p + 1) - 2\sqrt{p} \leq \#E(\mathbb{F}_p) \leq (p + 1) + 2\sqrt{p}.$$

The interval length for Hasse's theorem is $4\sqrt{p}$ so that if there is an $a \in E(\mathbb{F}_p)$ with order greater than $4\sqrt{p}$ then, as the order of a divides the order of $E(\mathbb{F}_p)$, some integer multiple of $|a|$ will be within the interval and this will be the only such multiple. Then the order of $E(\mathbb{F}_p)$ will be this multiple of $|a|$. The order of a given element may be calculated using Shanks Baby - Step Giant - Step method so that the remaining question is whether there exist elements of order greater than $4\sqrt{p}$ in $E(\mathbb{F}_p)$. The answer to this is in general no but this may be remedied by considering the quadratic twist of an elliptic curve E defined below.

Given a quadratic non - residue n, the quadratic twist of E given by $y^2 = x^3 + ax + b$ defined over K is denoted E' and is given by

$$y^2 = n^{-1}(x^3 + ax + b).$$

E and E' are not isomorphic over K but are isomorphic over $K(\sqrt{n})$. n^{-1} is a quadratic non - residue so that for $x \in \mathbb{F}_p$ such that $x^3 + ax + b$ is a quadratic residue, $n^{-1}(x^3 + ax + b)$ is a quadratic non - residue so that x affords two solutions on E and none on E'. Similarly, for x such that $x^3 + ax + b$ is a quadratic non - residue, $n^{-1}(x^3 + ax + b)$ is a quadratic residue and x then affords two solutions on E' and none on E. For x such that $x^3 + ax + b = 0$, $n^{-1}(x^3 + ax + b) = 0$ as well so that $(x, 0)$ is a solution on both E, E'. Thus for each $x \in \mathbb{F}_P$, there are two solutions among E and E', and each curve has a point at infinity as well so that

$$\#E(\mathbb{F}_p) + \#E'(\mathbb{F}_p) = 2p + 2.$$

Thus if the order of one of the two curves is known, then so is the order of the other.

Again, one of E, E' must have an element of order greater than $4\sqrt{p}$ in order for the Shanks - Mestre algorithm to work, and the following theorem states that this is so [1].

Theorem 14. For $p > 19$, one of E, E' has an element of order greater than $4\sqrt{p}$.

Thus, the above method may be implemented, and the idea of the Shanks - Mestre algorithm is as follows:

Begin with a point $a \in E(\mathbb{F}_p)$ and compute $|a|$ using Shanks Baby - Step Giant - Step method. If $|a| > 4\sqrt{p}$, deduce $\#E(\mathbb{F}_p)$. If $|a| < 4\sqrt{p}$, choose a point q on $E'(\mathbb{F}_p)$ and again, if $|q| > 4\sqrt{p}$, deduce $\#E'(\mathbb{F}_p)$ and if $|q| < 4\sqrt{p}$, restart with a different point $b \in E(\mathbb{F}_p)$.

The running time of this algorithm is $O(p^{1/4+\epsilon})$ for any $\epsilon > 0$, and should be used in place of the algorithm using the Legendre symbol as soon as p is greater than 100, say. Again, a full description of the algorithm may be found in appendix B, and further information on the Shanks - Mestre algorithm and Schoof's algorithm may be found in [1].

Chapter 2

Data Analysis and Methodology

Let $\pi(x)$ denote the number of primes less than x. The following describes the distribution of primes within the integers;

Theorem 15. Prime Number Theorem

$$\pi(x) \sim \frac{x}{\log(x)}.$$

Thus $\frac{1}{\log(x)}$ may be thought of as the proportion of primes less than x, that

$$\frac{\#\{p \mid p \text{ is prime}, p < x\}}{x} \sim \frac{1}{\log(x)}.$$

From the Sato - Tate conjecture,

$$\frac{2}{\pi} \int_{-1}^{-1+\frac{1}{2\sqrt{p}}} \sqrt{1-t^2}\, dt = \frac{2}{\pi} \int_{-1}^{-1+\frac{1}{2\sqrt{p}}} \left(\sqrt{2}(1-t)^{1/2} + O((1-t)^{3/2})\right) dt$$

$$= \frac{2}{\pi}\left(\frac{2\sqrt{2}}{3}\left(\frac{1}{2\sqrt{p}}\right)^{3/2}\right) + O(p^{-5/4})$$

$$= \frac{2}{3\pi}p^{-3/4} + O(p^{-5/4})$$

is the 'probablilty' that a prime is a champion prime. Then summing over all primes up to x,

ignoring the error term, and using the prime number theorem gives

$$\pi_E^{\text{Champ}}(x) \sim \frac{2}{3\pi} \sum_{p<x} p^{-3/4} \sim \frac{2}{3\pi} \int_2^x \frac{dt}{t^{3/4}\log(t)}.$$

Thus it is reasonable to use

$$\int_2^x \frac{dt}{t^{3/4}\log(t)}$$

as a linear predictor for $\pi_E^{\text{Champ}}(x)$, that is to suppose

$$\pi_E^{\text{Champ}}(x) = \beta_1 \int_2^x \frac{dt}{t^{3/4}\log(t)} + \beta_0$$

and to see whether β_1 is approximately $\frac{2}{3\pi}$ for various non - CM curves. The supposition that there is a single linear predictor $\int_2^x \frac{dt}{t^{3/4}\log(t)}$ for the outcome $\pi_E^{\text{Champ}}(x)$ is exactly the assumption for simple linear regression, the method used to help answer the above question.

2.1 Introduction to Data Analysis, Techniques Used

As stated above, the method used to perform the data analysis in the investigation is simple linear regression. In order to perform simple linear regression analysis, there must be a single predictor, X, and an outcome, Y, and an assumption that the outcome is a linear function of the predictor;

$$Y = \beta_1 X + \beta_0,$$

where β_1 and β_0 are the true slope and intercept of the linear function which predicts Y from X. However in order to account for inevitable deviations from the prediction of y_i from x_i, an error term is included in the model;

$$y_i = \beta_1 x_i + \beta_0 + \epsilon_i,$$

or

$$Y = \beta_1 X + \beta_0 + \epsilon,$$

where Y, ϵ are random variables which depend on one another.

In estimating the parameters β_1 and β_0, realizations X_0 of X and Y_0 of Y are used to produce a line estimating the linear prediction function. There are many ways to produce a line

estimating Y_0 from X_0, and the method used in this paper is least - squares linear regression. Least squares regression produces estimators $\hat{\beta}_1$, $\hat{\beta}_0$ of β_1, β_0 which minimize the quantity

$$\sum_i (y_i - (\tilde{\beta}_1 x_i + \tilde{\beta}_0))^2,$$

over all $\tilde{\beta}_1$ and $\tilde{\beta}_0$, where $y_i - (\tilde{\beta}_1 x_i + \tilde{\beta}_0)$ is called a residual. For more information on statistical analysis, see [6]. The $\hat{\beta}_1$ and $\hat{\beta}_0$ produced are then used as the coefficients of the predictor function

$$Y \approx \hat{\beta}_1 X + \hat{\beta}_0 = \hat{Y}$$

and the $\hat{\beta}_1$s produced in this way are what were compared in order to address whether the distributions are different between two non - CM elliptic curves.

In order to compare the $\hat{\beta}_1$s produced for different elliptic curves, for each elliptic curve a 95% confidence interval was computed for its associated $\hat{\beta}_1$. The interpretation associated to a 95% confidence interval for a parameter estimate $\hat{\mu}$ is that if a large number of 95% of confidence intervals were computed with various samples of the given data, 95% of these computed intervals would contain the true parameter μ. It is thus reasonable to expect that the computed confidence interval for a given $\hat{\beta}_1$ contains the true coefficient β_1 for the given elliptic curve associated to $\hat{\beta}_1$. These confidence intervals for the $\hat{\beta}_1$s are then compared to see if there are any which do not overlap.

Another method for comparing two coefficient estimates $\hat{\beta}_1$ of E, $\hat{\gamma_1}$ of E' is by using multilinear regression on the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$$

where X_2 is an indicator variable which is 0 for data associated to E and 1 associated to data from E', and where β_3 is called the cross term. When $X_2 = 0$ the model is

$$Y = \beta_0 + \beta_1 X_1$$

and when $X_2 = 1$ the model is

$$Y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1.$$

Note then that the first model is exactly the model for the linear prediction function of E already used to produce $\hat{\beta}_1$ and thus β_2 and β_3 compare the difference in the y-intercept and the slope of the model used for E to produce $\hat{\beta}_1$ and the model used for E' used to produce $\hat{\gamma}_1$. Then a 95% confidence interval is computed for $\hat{\beta}_3$ to see whether the confidence interval for $\hat{\beta}_3$ does not contain zero. If so then it would be likely that the parameter β_3 is non - zero, i.e. that the parameters β_1 and γ_1 are different.

2.2 An Example

To elucidate the methods outlined in the previous section, data for an example elliptic curve is computed. For the elliptic curve $E_{1,-5,-5,0,0}$, the following data was obtained - note that the x values are champion primes:

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{1,-5,-5,0,0}}(x)$	x
25.8885	1	34390831
28.2118	2	56781601
33.3809	3	149217361
37.0926	4	270705557
37.5984	5	292077259
38.9518	6	355979177
52.9303	7	1918202933
63.1168	8	4927943393

Table 2.1: Data for $E_{1,-5,-5,0,0}$

This data produces $\hat{\beta}_1 = 0.18282$, $\hat{\beta}_0 = -2.748$ from least squares regression, that

$$\pi^{\text{Champ}}_{E_{1,-5,-5,0,0}}(x) \approx 0.18282 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -2.748.$$

The 95% confidence interval for the slope is (0.13450, 0.23114).

The data for the elliptic curve $E_{0,0,0,9,53}$ is reproduced on the next page. This data produces

$\hat{\beta}_1 = 0.19706$, $\hat{\beta}_0 = -0.7304$ from least squares regression, that

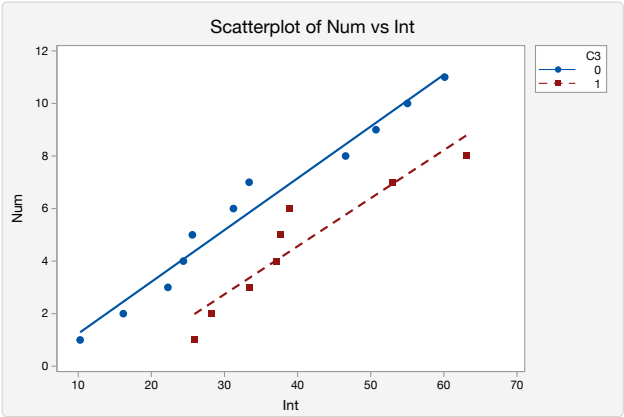
$$\pi^{\text{Champ}}_{E_{0,0,0,9,53}}(x) \approx 0.19706 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -0.7304.$$

The 95% confidence interval for the slope is (0.16654, 0.22758).

The confidence intervals for the slopes overlap so that it is possible that these two coefficients are the same. The 95% confidence interval for the cross term in the multilinear regression model is (-0.04292, 0.07139) which indicates that the true parameter β_3, which measures the cross term, may be 0, that there is not statistical significance between these two coefficients with the data provided. This is seen in a scatterplot of the two data, reproduced below - note that the linear prediction functions for both are roughly parallel.

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{0,0,0,9,53}}(x)$	x
10.2617	1	107791
16.1648	2	1988587
22.2642	3	14073491
24.3813	4	24155983
25.6068	5	32250949
31.2267	6	101968327
33.3706	7	148955167
46.5626	8	955017131
50.7152	9	1521637807
55.0259	10	2365283659
60.1147	11	3800917487

Table 2.2: Data for $E_{0,0,0,9,53}$



(a)

Equation

Num = −0.730356 + 0.197059 Int

Num = −2.74828 + 0.182823 Int

(b)

Figure 2.1: (a) Comparison of the regression lines for $E_{0,0,0,9,53}$, in blue and $E_{1,-5,-5,0,0}$, in red and (b) Equations of the regression lines for the two curves

2.3 A Number Theoretic Perspective

As noted above, the prime number theorem asymptotically gives the distribution of primes within the integers. Understanding the error of this asymptotic is of interest. Assuming the Riemann Hypothesis, the error is much less than $x^{1/2}\log(x)$, that is that

$$\left|\pi(x) - \frac{x}{\log(x)}\right| \ll x^{1/2}\log(x).$$

This is an upper bound and the error is often less than $x^{1/2}$.

From before, on average over all elliptic curves,

$$\pi_E^{\text{Champ}}(x) \sim \frac{8x^{1/4}}{3\pi\log(x)}$$

and from the above discussion, if $\pi_E^{\text{Champ}}(x)$ is not asymptotically curve dependent, it is reasonable to expect

$$\left|\pi_E^{\text{Champ}}(x) - \frac{8x^{1/4}}{3\pi\log(x)}\right| \ll x^{1/8}\log(x)$$

with the error term often being less than $x^{1/8}$. In this case it would then be reasonable to expect that for two elliptic curves E_1, E_2,

$$\left|\pi_{E_1}^{\text{Champ}}(x) - \pi_{E_2}^{\text{Champ}}(x)\right| < x^{1/8}.$$

Chapter 3

Conclusions and Discussion

3.1 Evidence for and against curve dependence of $\pi_E^{\mathbf{Champ}}(x)$

Within the set of curves tested during the investigation this paper discusses, $E_{0,0,0,17,32143}$ and $E_{1,-783/16,-783/16,0,0}$ were the curves with the highest and lowest predicted coefficient on the linear predictor function, respectively. In the interval from 2 to 10.8 billion, $E_{1,-783/16,-783/16,0,0}$ has 6 champion primes, has linear predictor function

$$\pi^{\text{Champ}}_{E_{1,-783/16,-783/16,0,0}}(x) \approx 0.0786 \int_2^x \frac{dt}{t^{3/4}\log(t)} + 0.2827,$$

and the 95% confidence interval for the slope is (0.02703, 0.13021). $E_{0,0,0,17,32143}$ has 24 champion primes, has linear predictor function

$$\pi^{\text{Champ}}_{E_{0,0,0,17,32143}}(x) \approx 0.3141 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -1.0958,$$

and the 95% confidence interval for the slope is (0.28873, 0.33950). The 95% confidence intervals for the slope do not overlap and suggest that the slopes are different; this is again seen in the 95% confidence interval for the cross term, (0.17800, 0.29299), which indicates that the cross term is likely not 0, that there is a statistical significance between the difference in the two slopes. Thus

there may be reason to suspect that the distribution functions

$$\pi^{\text{Champ}}_{E_{0,0,0,17,32143}}(x) \text{ and } \pi^{\text{Champ}}_{E_{1,-783/16,-783/16,0,0}}(x)$$

are different. This may be seen in a plot of the two regression lines, below.

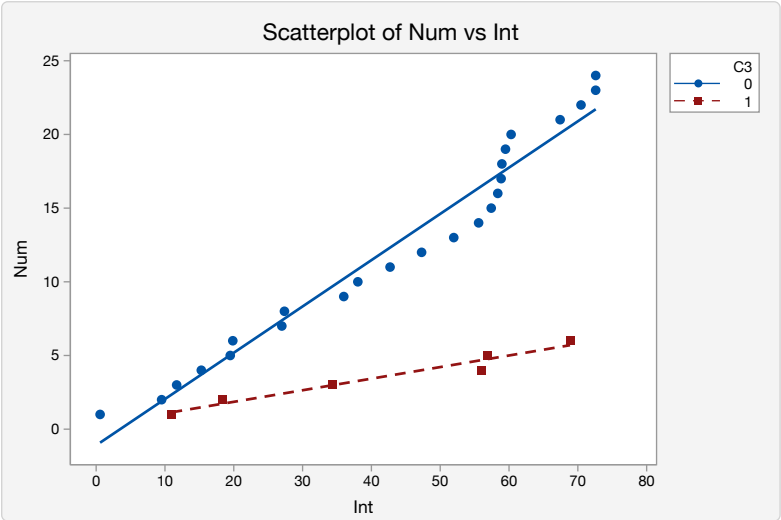


Figure 3.1: Comparison of the regression lines for $E_{0,0,0,17,32143}$ in blue, $E_{1,-783/16,-783/16,0,0}$ in red

As stated above, if $\pi^{\text{Champ}}_{E}(x)$ is asymptotically not curve dependent, it should be reasonable to expect

$$\left|\pi^{\text{Champ}}_{E_1}(x) - \pi^{\text{Champ}}_{E_1}(x)\right| < x^{1/8}$$

for two elliptic curves E_1 and E_2. The two curves in the previous paragraph, $E_{0,0,0,17,32143}$ and $E_{1,-783/16,-783/16,0,0}$, have the highest and lowest number of champion primes, respectively, within the curves checked for the investigation discussed in this paper. If $\pi^{\text{Champ}}_{E}(x)$ is not asymptotically curve dependent, then it would be reasonable to expect

$$\left|\pi^{\text{Champ}}_{E_{0,0,0,17,32143}}(10800000000) - \pi^{\text{Champ}}_{E_{1,-783/16,-783/16,0,0}}(10800000000)\right| < 10800000000^{1/8} \approx 17.955.$$

From above,

$$\pi^{\text{Champ}}_{E_{0,0,0,17,32143}}(10800000000) = 24$$

and

$$\pi^{\text{Champ}}_{E_{1,-783/16,-783/16,0,0}}(10800000000) = 6$$

so that

$$\left|\pi^{\text{Champ}}_{E_{0,0,0,17,32143}}(10800000000) - \pi^{\text{Champ}}_{E_{1,-783/16,-783/16,0,0}}(10800000000)\right| = 24 - 6 = 18.$$

Thus the difference in the number of champion primes is around the expected maximum of the error term if $\pi^{\text{Champ}}_E(x)$ is asymptotically not curve dependent, and from this perspective there may be reason to suspect that $\pi^{\text{Champ}}_E(x)$ is asymptotically not curve dependent.

3.2 Constraints and Potential Directions

In the previous section it was seen that the two methods used to provide evidence for or against an asymptotically curve dependent distribution function $\pi^{\text{Champ}}_E(x)$ gives inconclusive results - the data analysis provides evidence that $\pi^{\text{Champ}}_E(x)$ may asymptotically be curve dependent and the number theoretic perspective provides evidence that $\pi^{\text{Champ}}_E(x)$ may asymptotically not be curve dependent. This conflicting evidence may be partially accounted for by the scarcity of champion primes; up to 5.4 billion, the greatest number of champion primes seen among the curves studied in this investigation is 20, and checking this curve for primes up to 10.8 billion produced 24. This is reinforced by the fact that the predictor function $\int_2^x \frac{dt}{t^{3/4}\log(t)}$ begins to flatten for large x;

$$\int_2^{5400000000} \frac{dt}{t^{3/4}\log(t)} \approx 64.21308,$$

while

$$\int_{5400000000}^{10400000000} \frac{dt}{t^{3/4}\log(t)} \approx 9.012350.$$

Thus new methods for investigating this problem may be needed.

Appendices

Appendix A Overview of Code and Implementation

In order to compute the champion primes for a given elliptic curve, the code below was used. As can be seen, 7 numbers are input to the function. The first five of these are the coefficients of the elliptic curve and for each prime p between the last two inputs, the curve is reduced modulo p and $\#E(\mathbb{F}_p)$ is computed to see whether p is a champion or trailing prime for E. J referenced in the code is the job number, and this along with the five lines at the top of the code are related to the fact that the code was parallelized and run on a remote computing cluster called the Palmetto cluster. The line at the bottom would be updated to have inputs corresponding the elliptic curve of interest. The conductor and torsion subgroup are for $E(\mathbb{Q})$ and do not change as m, n change so it is only needed once, so was output for $J = 0$.

The code below is the text of a file used on the Palmetto cluster to compute champion primes for a given elliptic curve E specified at the bottom of the code. Following the program is a .pbs file which is used to submit jobs to the cluster. This file was saved in the home directory and was moved to a specific elliptic curve E's directory before the job was submitted. The outputs at the top are for specifying attributes about the job. The second line from the top specifies the amount of memory requested for the job, the amount of time requested for the job, and the number of c.p.u.s and nodes used for each job. The fourth line from the top specifies that 12 jobs will be submitted, and that J is used to reference a specific job. The fifth line from the top says to load sage onto the node on which a job is being run, and the sixth line says to run the sage code in a specific elliptic curve E's directory and to output the log files and output files in the same directory, labeled by job number.

For each non - CM elliptic curve, 36 jobs were submitted using three separate .pbs files, submitting 12 jobs per .pbs file, and the jobs were run on the Palmetto cluster over the course of a day. For these non - CM curves, the interval over which champion primes were checked is 3 to 5.4 billion - an interval length of 150 million for each job. Note that this code is parallelized in the sense that separate intervals are being computed at the same time and compiled into a total list of champion primes and trailing primes after the program terminates and outputs the lists for their respective intevals.

```
import sys
J = int(sys.argv[1])
print J
tm = 150000000*(J)+2
tn = 150000000*(J+1)+2
def evala4(a,b,c,d,e,m,n):
    E = EllipticCurve([a,b,c,d,e])
    cp = [0]
    tp = [0]
    cpc = 0
    tpc = 0
    ch = []
    tr = []
    l=0
    k = 0
    if J == 0:
        print 'Torsion is', E.torsion_subgroup()
        print 'Conductor is', E.conductor()
        print 'Has CM?', E.has_cm()
    P=Primes()
    i = P.next(m)
    while i<n:
        if E.discriminant() % i != 0:
            l = (i+1) - E.Np(i)
            if l == -floor(2*sqrt(i)):
                k = i
                cpc = cpc+1
                cp.append(i)
                ch.append(cpc)
                ch.append(RR(Ei(1/4*log(i)) - Ei(1/4*log(2))))
            if l == floor(2*sqrt(i)):
                k = i
                tpc = tpc+1
                tp.append(i)
                tr.append(tpc)
                tr.append(RR(Ei(1/4*log(i)) - Ei(1/4*log(2))))
        i = P.next(i)
    print ch
    print tr
    print  i, cpc, tpc, cp, tp, RR(Ei(1/4*log(i)) - Ei(1/4*log(2))), tm, tn
    return()

evala4(0,0,0,0,0, tm, tn)
```

Figure 2: Program to find Champion Primes; 'round1.sage'

```
#PBS -N oe
#PBS -l select=1:ncpus=1:mem=10gb,walltime=04:30:00
#PBS -j oe
#PBS -J 0-11


module add sage/6.1.1

sage /home/arhahn/E's directory/round1.sage "$PBS_ARRAY_INDEX"
> /home/arhahn/E's directory/${PBS_ARRAY_INDEX}.log
```

Figure 3: .pbs file

```
0
Torsion is Torsion Subgroup isomorphic to Trivial group associated to
the Elliptic Curve defined by y^2 = x^3 + 17*x + 32143 over Rational Field
Conductor is 4057552840
Has CM? False
[1, 0.574192777601945, 2, 9.51467769075006, 3, 11.6992178405152, 4,
15.2800752480566, 5, 19.4859558420902, 6, 19.8564097592252, 7,
26.9813555205544, 8, 27.3621757937467]
[1, 2.16253107742731, 2, 4.00324090918191, 3, 7.36165991957638, 4,
23.0261998438208, 5, 25.9052015857281, 6, 26.2093991979151, 7,
27.3966766098115, 8, 32.5364723045627]
150000029 8 8 [0, 3, 65557, 253633, 1397579, 6294649, 7058647, 43803259,
47532977]
[0, 19, 271, 12011, 17204711, 34521061, 36966049, 47883139, 128963587]
33.4116614344546
2 150000002
```

Figure 4: Sample output for job 0 of $E_{0,0,0,17,32143}$

The 0 at the top of the sample output for $E_{0,0,0,17,32143}$ is the job number. The first array output after the text is the count of champion primes and the integral values at which the champion primes occurred. Similarly, the second array is the count of the trailing primes and the integral values at which the trailing primes occurred. Below that is the final output of the job which is the last prime computed, the number of champion primes, the number of trailing primes, the champion primes, the trailing primes, the integral value of the last prime computed, and the interval over which champion primes and trailing primes were computed.

Appendix B Shanks - Mestre Algorithm

A description of the Shanks - Mestre algorithm is given in the chapter on elliptic curves and motivation. A full description of the algorithm may be found in [1] and the algorithm itself is reproduced below.

Algorithm 16. (Shanks - Mestre) Given an elliptic curve E over $\mathbb{F}_p$ with $p \geq 13$ by a Weierstraß equation $y^2 = x^3 + ax + b$, this algorithm computes the a_p such that $\#E(\mathbb{F}_p) = p + 1 - a_p$

1. (Initialize) Set $x \leftarrow -1, A \leftarrow 0, B \leftarrow 1, k_1 = 0$.

2. (Get next point) [Here we have $\#E(\mathbb{F}_p) \equiv A \pmod{B}$] Repeat $x \leftarrow x + 1, d \leftarrow x^3 + ax + b$, $k \leftarrow \left(\frac{d}{p}\right)$ until $k \neq 0$ and $k \neq k_1$. Set $k_1 \leftarrow k$. Finally, if $k_1 = -1$ set $A_1 \leftarrow 2p + 2 - A \bmod B$ else set $A_1 \leftarrow A$.

3. (Find multiple of the order of a point) Let m be the smallest integer such that $m > p+1-2\sqrt{p}$ and $m \equiv A_1 \pmod{B}$. Using Shank's baby - step ginat - step strategy, find an integer n such that $m \leq n < p + 1 + 2\sqrt{p}$, $n \equiv m \pmod{B}$ and such that $n \cdot (xd, d^2) = 0$ on the curve $Y^2 = X^3 + ad^2X + bd^3$ [note that this will be isomorphic to the curve E or E' according to the sign of k_1].

4. (Find order) Factor n, and deduce from this the exact order h of the point (xd, d^2).

5. (Finished?) Using for instance the Chinese remainder algorithm, find the smallest integer h' which is a multiple of h and such that $h' \equiv A_1 \pmod{B}$. If $h' < 4\sqrt{p}$ set $B \leftarrow LCM(B, h)$, then $A \leftarrow h' \bmod B$ if $k_1 = 1$, $A \leftarrow 2p + 2 - h' \bmod B$ if $k_1 = -1$, and go to step 2.

6. (Compute a_p) Let N be the unique multiple of h' such that $p + 1 - 2\sqrt{p} < N < p + 1 + 2\sqrt{p}$. Output $a_p = p + 1 - k_1 N$ and terminate the algorithm.

The running time of this algorithm is $O(p^{1/4+\epsilon})$ for any $\epsilon > 0$, and should be used in place of the algorithm using the Legendre symbol as soon as p is greater than 100, say.

Appendix C Data

The data for the curves is listed below. [3, 10 800 000 000] is the interval over which data was collected for $E_{1,-783/16,-783/16,0,0}$ and $E_{0,0,0,0,17,32143}$, while data was collected over the interval [3, 5 400 000 000] for the remaining non - CM curves.

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{-11,-12,-12,0,0}}(x)$	x
1.1080	1	5
3.4672	2	127
13.1328	3	534407
17.4562	4	3205459
29.7321	5	76920149
35.1103	6	198673289
41.0837	7	478917631
48.5794	8	1204047641
49.5053	9	1334473139
50.3917	10	1469674357
55.4818	11	2472706267
56.9672	12	2849981089

Table 1: Data for $E_{-11,-12,-12,0,0}$

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{-10,-132,-132,0,0}}(x)$	x
12.0906	1	313879
20.1012	2	7603601
21.0859	3	10153007
25.2809	4	29911649
32.2827	5	123330811
33.5624	6	153887501
35.7244	7	219103883
40.0545	8	415925339
44.9011	9	782379991
44.9264	10	784811219
48.8666	11	1243361851
54.5661	12	2260739933
56.1988	13	2649510169
62.0880	14	4515327367
62.6979	15	4756472813

Table 2: Data for $E_{-10,-132,-132,0,0}$

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{1,-\frac{783}{16},-\frac{783}{16},0,0}}(x)$	x
10.9423	1	164117
18.3679	2	4385987
34.4034	3	177079897
56.0223	4	2605088039
56.9270	5	2839213823
68.8632	6	7817589469

Table 3: Data for $E_{1,-783/16,-783/16,0,0}$

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{-41,-294,-294,0,0}}(x)$	x
2.9650	1	61
6.9293	2	8059
16.9596	3	2680819
24.158	4	22876213
24.8463	5	27006757
30.2845	6	85523359
30.8003	7	94237243
30.8737	8	95533027
35.9661	9	227591263
44.8528	10	777764063

Table 4: Data for $E_{-41,-294,-294,0,0}$

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{0,0,0,9,53}}(x)$	x
10.2617	1	107791
16.1648	2	1988587
22.2642	3	14073491
24.3813	4	24155983
25.6068	5	32250949
31.2267	6	101968327
33.3706	7	148955167
46.5626	8	955017131
50.7152	9	1521637807
55.0259	10	2365283659
60.1147	11	3800917487

Table 5: Data for $E_{0,0,0,9,53}$

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{1,-5,-5,0,0}}(x)$	x
25.8885	1	34390831
28.2118	2	56781601
33.3809	3	149217361
37.0926	4	270705557
37.5984	5	292077259
38.9518	6	355979177
52.9303	7	1918202933
63.1168	8	4927943393

Table 6: Data for $E_{1,-5,-5,0,0}$

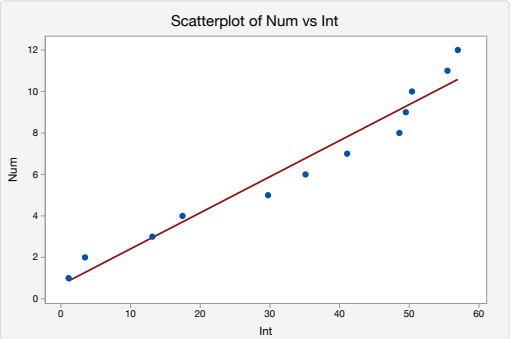
$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{1,5,5,0,0}}(x)$	x
2.8694	1	53
4.0032	2	271
13.8445	3	748589
15.4878	4	1521301
17.9734	5	3837877
21.4504	6	11256281
24.5401	7	25100503
30.3190	8	86083651
35.8625	9	223923187
39.9412	10	409417483
40.8132	11	461666771
47.3297	12	1044340391
49.8966	13	1392879821
50.0370	14	1414323133
51.5088	15	1655366927
58.9016	16	3408498251

Table 7: Data for $E_{1,5,5,0,0}$

$\int_2^x \frac{dt}{t^{3/4}\log(t)}$	$\pi^{\text{Champ}}_{E_{0,0,0,17,32143}}(x)$	x
0.5742	1	3
9.5147	2	65557
11.6992	3	253633
15.2801	4	1397579
19.4860	5	6294649
19.8564	6	7058647
26.9814	7	43803259
27.3622	8	47532977
35.9845	9	228247183
38.0291	10	311320951
42.7120	11	593960053
47.2944	12	1040084693
51.9684	13	1736990201
55.5820	14	2496827657
57.4126	15	2971611029
58.3641	16	3245274509
58.8510	17	3392838103
58.9624	18	3427366409
59.4838	19	3592601267
60.3018	20	3864533801
67.4245	21	6992872847
70.4609	22	8822079743
72.5917	23	10318227923
72.5989	24	10323609509

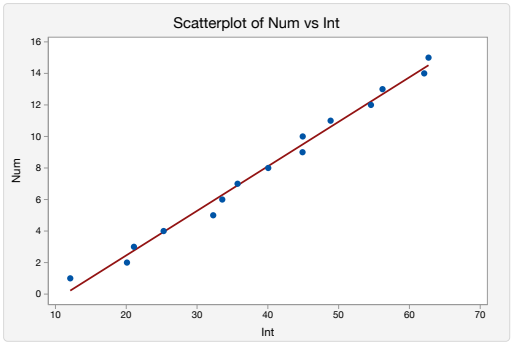
Table 8: Data for $E_{0,0,0,17,32143}$

The following are the regression lines for each set of the above data.



(a) $0.1740 \int_2^x \frac{dt}{t^{3/4}\log(t)} + 0.6709$;

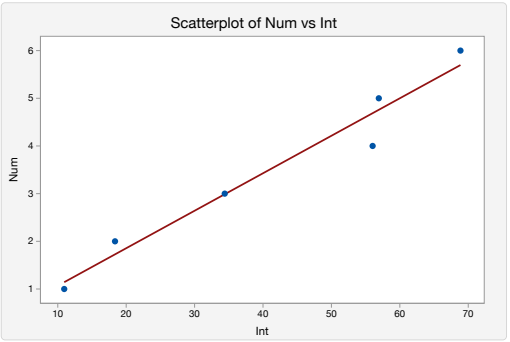
95% CI for slope: $(0.14846, 0.19954)$



(b) $0.2825 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -3.1946$;

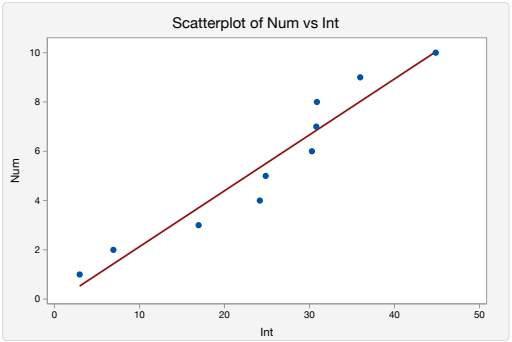
95% CI for slope: $(0.26570, 0.29928)$

Figure 5: (a) Regression line for $E_{-11,-12,-12,0,0}$ and confidence interval for the slope, and (b) regression line for $E_{-10,-132,-132,0,0}$ and confidence interval for the slope



(a) $0.0786 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -0.2827$;

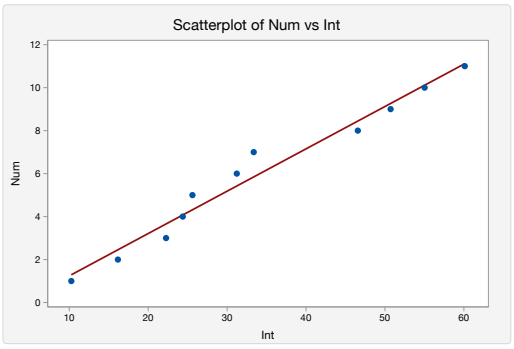
95% CI for slope: $(0.02703, 0.13021)$



(b) $0.2269 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -0.1419$;

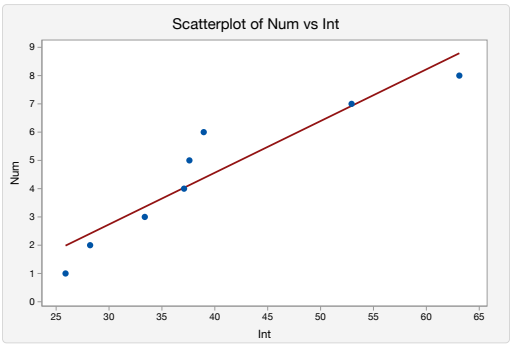
95% CI for slope: $(0.17973, 0.27410)$

Figure 6: (a) Regression line for $E_{1,-783/16,-783/16,0,0}$ and confidence interval for the slope, and (b) regression line for $E_{-41,-41,-294,0,0}$ and confidence interval for the slope



(a) $0.19706 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -0.7304$;

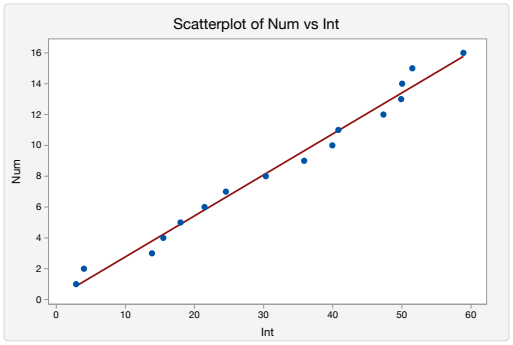
95% CI for slope: $(0.16654, 0.22758)$



(b) $0.18282 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -2.748$;

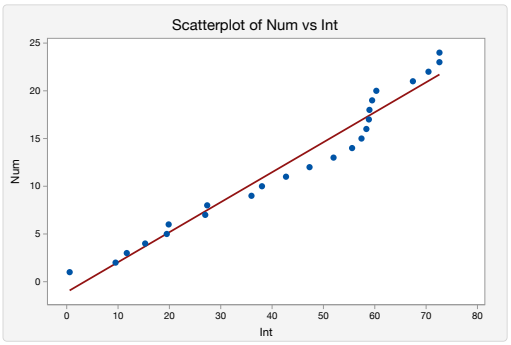
95% CI for slope: $(0.13450, 0.23114)$

Figure 7: (a) Regression line for $E_{0,0,0,9,53}$ and confidence interval for the slope, and (b) regression line for $E_{1,-5,-5,0,0}$ and confidence interval for the slope



(a) $0.2659 \int_2^x \frac{dt}{t^{3/4}\log(t)} + 0.1113$;

95% CI for slope: $(0.24814, 0.28365)$



(b) $0.3141 \int_2^x \frac{dt}{t^{3/4}\log(t)} + -1.0958$;

95% CI for slope: $(0.28873, 0.33950)$

Figure 8: (a) Regression line for $E_{1,5,5,0,0}$ and confidence interval for the slope, and (b) regression line for $E_{0,0,0,17,32143}$ and confidence interval for the slope

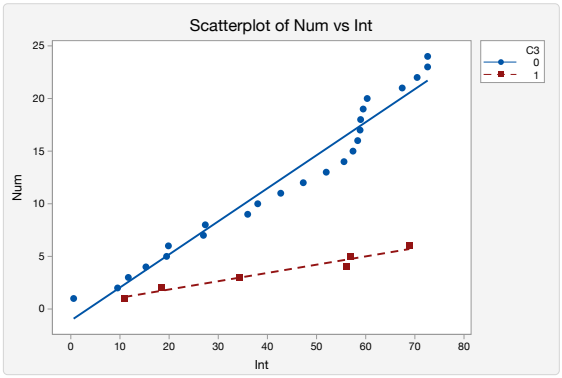


Figure 9: Comparison of the regression lines for $E_{0,0,0,17,32143}$ in blue, $E_{1,-783/16,-783/16,0,0}$ in red

Bibliography

[1] Henri Cohen. *A Course in Computational Algebraic Number Theory.* Springer-Verlag, 1993.

[2] D.A. Cox. *Primes of the form x^2+ny^2: Fermat, Class Field Theory, and Complex Multiplication.* Wiley, 2011.

[3] Luke Giberson and Kevin James. An average asymptotic for the number of extremal primes of elliptic curves. *Acta Arithmetica*, 2018.

[4] Kevin James. Average frobenius distributions for elliptic curves with 3-torsion. *Journal of Number Theory*, 2004.

[5] Kevin James and Paul Pollack. Extremal primes for elliptic curves with complex multiplication. *Journal of Number Theory*, 2017.

[6] R. Lyman Ott and Michael Longnecker. *An Introduction to Statistical Methods and Data Analysis.* Duxbury, 2001.

[7] Joseph Silverman and John Tate. *Rational Points on Elliptic Curves.* Springer, 2nd edition, 2015.